

Detección de la roya en café Caturra, a partir de datos de una red de sensores inalámbrica, utilizando árboles de decisión

Autores: Juan G. Cárcamo, Alberto Restrepo, Mauricio Toro

Consideraciones iniciales



Trabajo en
parejas



Puntuación extra si
lo escriben y
sustentan en
inglés



Usar **plantilla
ACM**



Entregar informe
en **PDF** y código
en **GIT**



Informe
**máximo en 4
páginas**



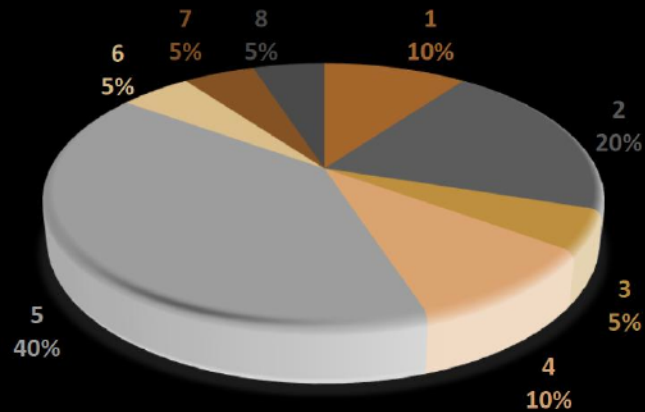
Detalles completos
en “**Guía para la
realización del
proyecto final**”

DOCENTE **MAURICIO TORO BERMÚDEZ**

Teléfono: (+57) (4) 261 95 00 Ext. 9473. Oficina: 19 - 627

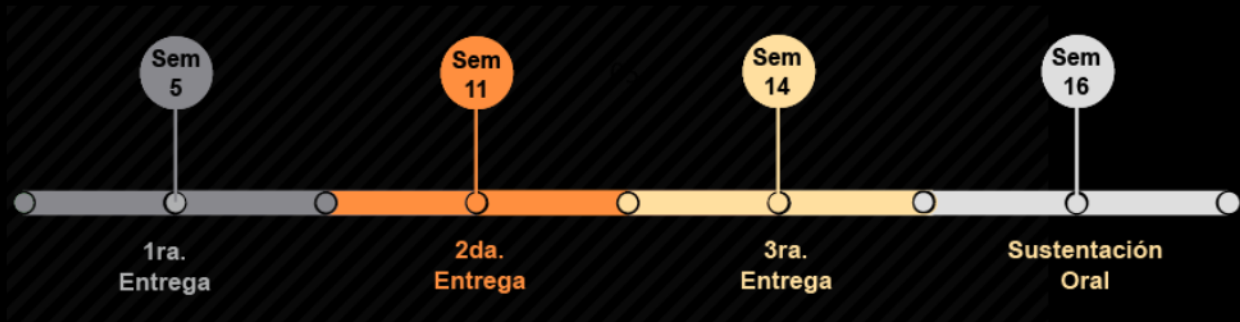
Correo: mtorobe@eafit.edu.co

Criterios de evaluación para el proyecto



- 1. Alternativas de Solución
- 2. Complejidad de Operaciones
- 3. Criterios de Forma para Código
- 4. Criterios de Fondo para Código
- 5. Criterios de Diseño Estructura de Datos
- 6. Informe Final
- 7. Progreso Gradual
- 8. Diapositivas

Tiempos de entrega en semanas académicas



Rúbricas de calificación

Lean la Sección 9 de la “*Guía para la realización del proyecto final de Estructuras de Datos 1*”

Intercambio de archivos



1. Motivación

Los tres principales productores de café en el mundo son Brasil (con 43.2 M de sacos al año), Vietnam (con 27.5M de sacos al año) y Colombia (con 13.5M de sacos al año). **En Colombia, el café es nuestra principal exportación agrícola y, aproximadamente, 563.000 familias dependen de él. La plaga de la roya es el principal problema fitosanitario que afecta al café.** El problema es agravado porque se hace un diagnóstico muy tarde. Esto hace que su control sea difícil y hay, inevitablemente, altas pérdidas. Aunque existen diversas variedades de café que son más resistentes a la roya, la variedad caturra (*coffea arabica*), que es la de exportación, es de las más susceptibles a esta plaga.

Una solución a este problema es hacer un monitoreo constante de las plantas de café utilizando redes de sensores inalámbricas. Investigadores de la Universidad Eafit desarrollaron un invernadero capaz de monitorear múltiples variables físico químicas que están asociadas a la aparición de la roya.



Gráfica 1. Invernadero para el monitoreo de cultivos de café desarrollado en la Universidad Eafit. Éste se encuentra contiguo al instituto del plástico en el campus de la universidad.

2. Problema

De acuerdo a la motivación, el problema que tenemos es **diseñar un algoritmo --basado en árboles de decisión-- para predecir si un lote de café caturra está infectado con la roya o no**. En particular, las variables fisicoquímicas que tenemos a disposición son: iluminación, temperatura ambiental, humedad del suelo, humedad ambiental, temperatura del suelo, ph del suelo. Con estas variables, el objetivo es crear un modelo que pueda predecir si un lote de café caturra, en un instante de tiempo dado, tiene la enfermedad de la roya. Además de las variables fisicoquímicas, para cada muestreo, contamos con la etiqueta de un experto que clasificó si ese lote de café tiene o no tiene la enfermedad de la roya.

DOCENTE **MAURICIO TORO BERMÚDEZ**

Teléfono: (+57) (4) 261 95 00 Ext. 9473. Oficina: 19 - 627

Correo: mtorobe@eafit.edu.co

3. Entradas y Salidas esperadas

El algoritmo debe recibir dos tipos de entradas. La primera entrada es un conjunto de datos, en formato archivo separado por comas (CSV), con el que se debe generar un árbol de decisión. Como un ejemplo, así vienen los datos organizados:

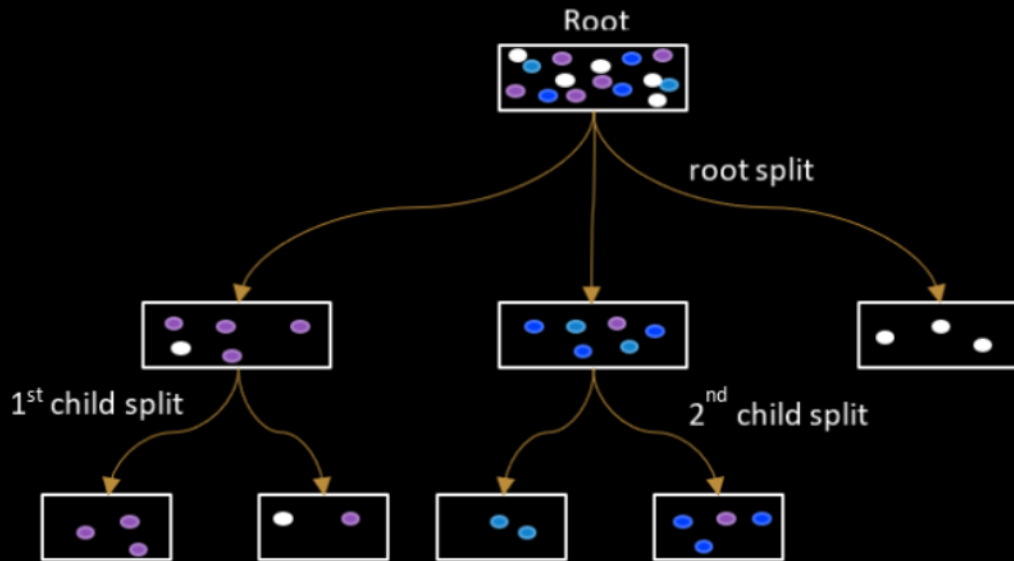
ph	soil_temperature	soil_moisture	illuminance	soil_temperature	env_humidity	label
6.41	21	22.1	1609	20	99	yes
4.52	21.75	18.84	3566	26	99	yes
5.85	23	59.06	568	27	40	no
7.42	23.25	44.57	2020	31	53	no

La segunda entrada es un nuevo registro. Ante la llegada de un nuevo registro, el algoritmo debe determinar la etiqueta (*label*); es decir, si hay o no presencia de la roya. Además, debe explicar por qué se tomó esa decisión. Como un ejemplo, si llega una nueva entrada como esta:

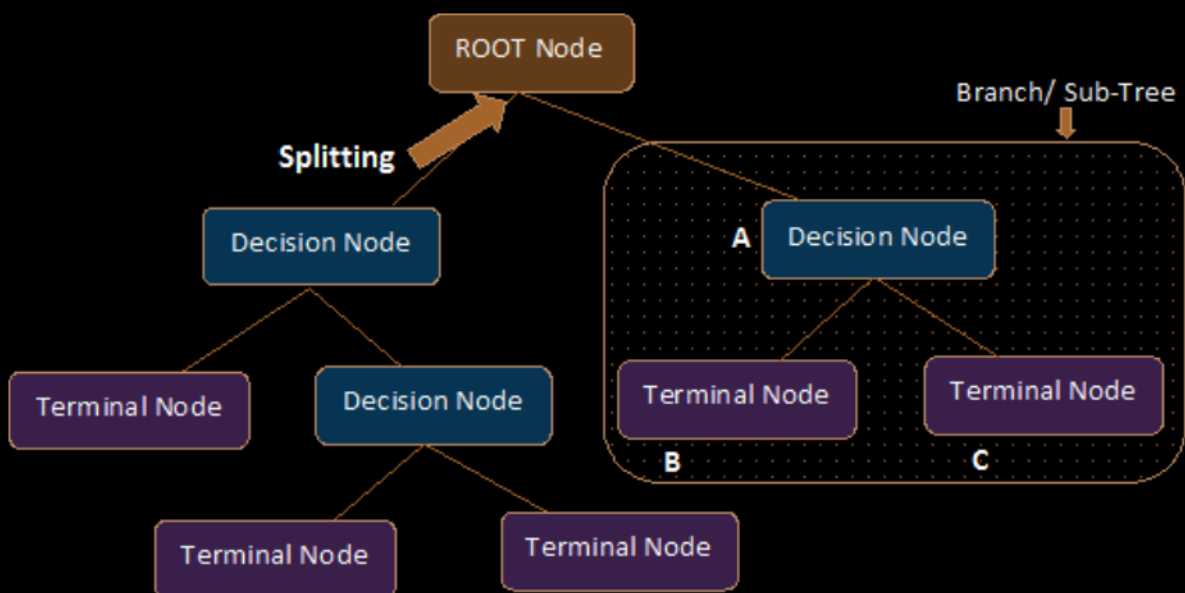
ph	soil_temperature	soil_moisture	illuminance	soil_temperature	env_humidity
7.41	20	22.1	1609	23	80

El algoritmo puede decir: La etiqueta es *yes* y la razón es que, según el árbol de decisión que ya se construyó con los datos de la primera entrada, como hay un *ph* mayor a 5, una temperatura del suelo menor a 21, y una humedad mayor a 75, hay presencia de la roya con una probabilidad del 73%.

Como un ejemplo, en la Gráfica 2 se muestra cómo las decisiones que se toman en un árbol de decisión separan el conjunto de datos en subconjuntos donde son más homogéneas las clases, es decir, donde están más diferenciados los registros en los que hay roya de los registros en los que no hay roya. Como otro ejemplo, la Gráfica 3 muestra cómo es la estructura de un árbol de decisión: Hay un nodo raíz, cuyos hijos son nodos de decisión y las hojas del árbol son nodos terminales.



Gráfica 2. Ejemplo del funcionamiento de un árbol de decisión



Note:- A is parent node of B and C.

Gráfica 3. Ejemplo de la estructura de un árbol de decisión

DOCENTE **MAURICIO TORO BERMÚDEZ**

Teléfono: (+57) (4) 261 95 00 Ext. 9473. Oficina: 19 - 627

Correo: mtorobe@eafit.edu.co

5. Algunos problemas relacionados

Para obtener información similar al problema aquí planteado, se sugiere ver los siguientes problemas relacionados:

a) Estructuras de datos y algoritmos relacionados

<https://www.youtube.com/watch?v=LDRbO9a6XPU>

https://es.wikipedia.org/wiki/Algoritmo_ID3

<https://es.wikipedia.org/wiki/C4.5>

https://en.wikipedia.org/wiki/Chi-square_automatic_interaction_detection

b) Tecnologías relacionadas

Caracol Radio. Universidad Eafit crea dispositivo para detectar en tiempo real la roya. Disponible en:

https://caracol.com.co/emisora/2019/07/08/medellin/1562601277_718794.html

A. Chemura, O. Mutanga, M. Sibanda, and P. Chidoko, "Machine learning prediction of coffee rust severity on leaves using spectroradiometer data," *Tropical Plant Pathology*, vol. 43, no. 2, pp. 117–127, 2018.

A. Chemura, O. Mutanga, and T. Dube, "Remote sensing leaf water stress in coffee (*coffea arabica*) using secondary effects of water absorption and random forests," *Physics and Chemistry of the Earth, Parts A/B/C*, vol. 100, pp. 317–324, 2017.

E. Lasso and J. C. Corrales, "Towards an alert system for coffee diseases and pests in a smart farming approach based on semi-supervised learning and graph similarity," in *International Conference of ICT for Adapting Agriculture to Climate Change*. Springer, 2017, pp. 111–123.